Super additive phonological similarity as constraint conjunction

Stephanie S Shih
University of California, Merced
shih@ucmerced.edu

1 INTRODUCTION

- **Phonological ganging** = type of cumulativity in constraint interaction wherein one strong constraint C_1 can be overtaken by two weaker constraints C_2 and C_3 together but not by C_2 or C_3 independently (e.g., Jäger & Rosenbach 2006).

(1) Ganging via (local) constraint conjunction in strict ranking OT
(e.g., Smolensky 1993, 2006; Baković 2000; Ito & Mester 2003; cf. Crowhurst & Hewitt 1997)

a.

	Input$_1$	C_2 & C_3	C_1	C_2	C_3
☞	Winner$_1$			-1	
	Loser$_1$		-1		

b.

	Input$_2$	C_2 & C_3	C_1	C_2	C_3
	Loser$_2$	-1		-1	-1
☞	Winner$_2$		-1		

- In weighted constraint Harmonic Grammar (HG) (e.g., Legendre et al. 1990; Smolensky & Legendre 2006):[1] argument that ganging can be achieved instead via cumulative constraint addition, without recourse to constraint conjunction (e.g., Farris-Trimble 2008; Potts et al. 2010; Pater 2015).

(2) Ganging via cumulative constraint addition in Harmonic Grammar
(e.g., Legendre, Sorace & Smolensky 2006; Potts et al. 2010)

a.

	weight	3	2	2	$\mathcal{H}$
	Input$_1$	C_1	C_2	C_3	
☞	Winner$_1$		-1		-2
	Loser$_1$	-1			-3

b.

	weight	3	2	2	$\mathcal{H}$
	Input$_2$	C_1	C_2	C_3	
	Loser$_2$		-1	-1	-4
☞	Winner$_2$	-1			-3

1.1 This talk

- **Proposal**: Additive constraint interaction in HG does not automatically preclude constraint conjunction. Cumulativity from additive constraint weights + cumulativity from constraint conjunction in a weighted constraint world = **super-cumulativity**. (aka: "super-additivity"; see e.g., Albright 2009; Green & Davis 2014)
- **Weighted constraint conjunction (WCC)** = each conjoined constraint C_1*C_2 also receives an independent weight, *above and beyond* the cumulative additive effect of the singular constraints C_1 and C_2.

(3) Super-cumulativity via Weighted Constraint Conjunction

a.

	weight	2	3	2	2	$\mathcal{H}$
	Input$_1$	C_2 & C_3	C_1	C_2	C_3	
☞	Winner$_1$			-1		-2
	Loser$_1$		-1			-3

b.

	weight	2	3	2	2	$\mathcal{H}$
	Input$_2$	C_2 & C_3	C_1	C_2	C_3	
	Loser$_2$	-1		-1	-1	-6
☞	Winner$_2$		-1			-3

[1] $\mathcal{H}\text{armony}(x, i) = \sum_{k=1}^{N} w_k C_i(x) = \boldsymbol{w_k} \cdot \boldsymbol{C_i}$; where x = candidate for input i, w_k= weight of constraint C_i, $C_i(x)$ = number of violations of C_i that x incurs, and N = vector of constraints (C_{i1} ... C_{iN}).

- Advantage of weighted constraint conjunction: more explanatory power in modeling patterns found in (variable) natural language data, where there are potential super-cumulative effects.

  ↳ Evidence presented today: Dioula d'Odienné tone harmony.
  ↳ Use of of information-theoretic model selection and comparison methodology to assess the contribution of weighted constraint conjunction to the grammar.
    – maximizes predictive accuracy while penalizing loss of restrictiveness (i.e., increased model complexity) in the addition of conjoined constraints (see also Wilson & Obdeyn 2009).

⇨ Significant loss of information and explanatory power when we *a priori* restrict our grammars against constraint conjunction, without thorough, quantitative assessments of their viability against noisy natural language data.

2 DATA

- Dioula d'Odienné (Mande, Côte d'Ivoire) (henceforth, Dioula)
- data = nouns from Braconnier & Diaby's lexicon (1982) (n = 1194)

- TYPE 1 lexical items: morphological H(igh) tone for definiteness only appears on the final vowel of the root.[2]
- TYPE 2 items: definite H tone triggers regressive H tone harmony on the final and penultimate syllables.[3]

		indefinite	*definite*			
(4)	a.	fòdà	fòd**á**	'season'	**TYPE 1**	L L → L H
	b.	brìsà	brìs**á**	'bush'		
	c.	sɛ̀bɛ̀	sɛ̀b**ɛ́**	'paper'		
	d.	hàmì	hàm**í**	'concern'		
(5)	a.	kùnà	**kúná**	'leprosy'	**TYPE 2**	L L → H H
	b.	tùrù	**túrú**	'oil'		
	c.	bègì	**bégí**	'white cotton cloth'		
	d.	bìlì	**bílí**	'flagstone terrace'		
	e.	mèlì	**mélí**	'worm'		
	f.	sànầ	**sánấ**	'tree'		

- Distinction between TYPE 1 and TYPE 2 nouns is predictable on at least three factors (Braconnier 1982; Shih 2013):
  1. sonority of the final intervocalic consonant (C_f);
  2. place identity of the final two vowels; and
  3. nasality of the final two vowels and C_f.

---

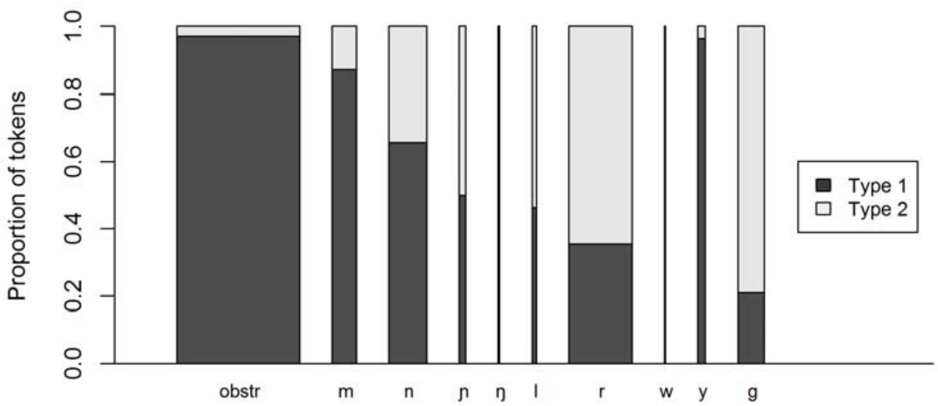
[2] Braconnier also reports that this pattern can be triggered by a H tone in the following word but provides no indication of the systematicity of that particular environment in triggering tone changes. Thus, this talk will focus on the case triggered by the definite suffix because evidence is more readily available in the lexicon. Additionally, this definite/indefinite alternation is characteristic of several Mande languages.

[3] In underlyingly high roots, the tone pattern is more complex: tone changes involve contour formation or LH spread. The basic pattern of tone change boundedness on the final two syllables versus the final syllable remains the same and can be modeled, with the addition of ANTI-HOMOPHONY constraints, under the same ABC analysis presented here. (for summaries, see Braconnier 1982; Shih 2013)

2.1 Sonority similarity

- Sonority effect scales quantitatively with increasing sonority within the sonorant series (Shih, ms.).
- Increase in sonority → increased likelihood of tone agreement (TYPE 2).

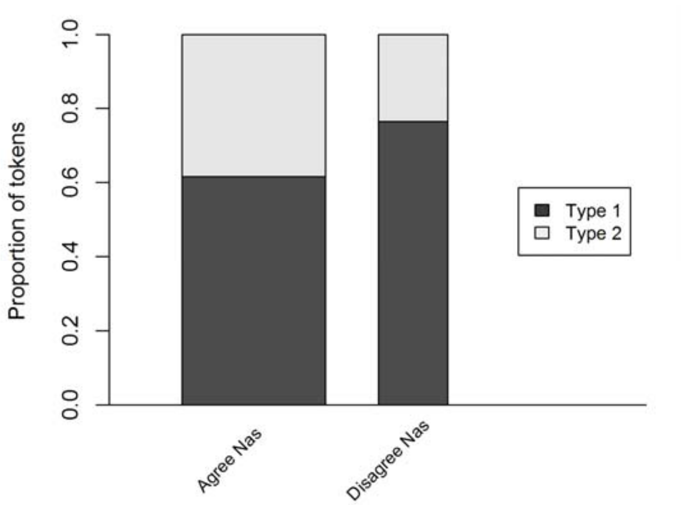
(6) C_fs by tone type[4]



2.2 Nasality similarity

- Words in which V_f and C_f agree in nasality or orality → increased likelihood of tone agreement (TYPE 2) (e.g., 5f).

(7) Nasality agreement by tone type

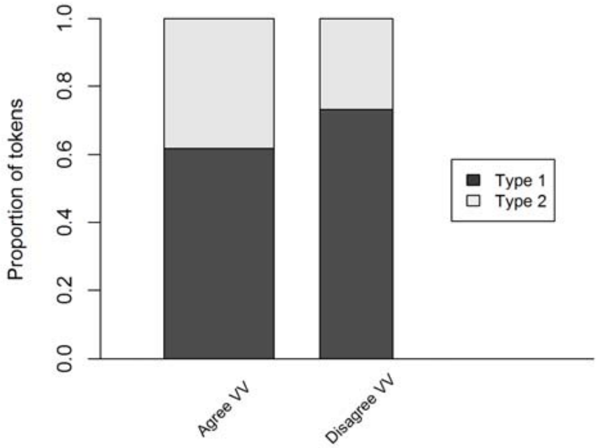


[4] Note: [g] ~ [ɣ] behaves phonotactically like a sonorant in Dioula.

2.3 V…V Identity

- Long distance vowel…vowel featural identity → increased likelihood of tone agreement (TYPE 2) (e.g., 5b).

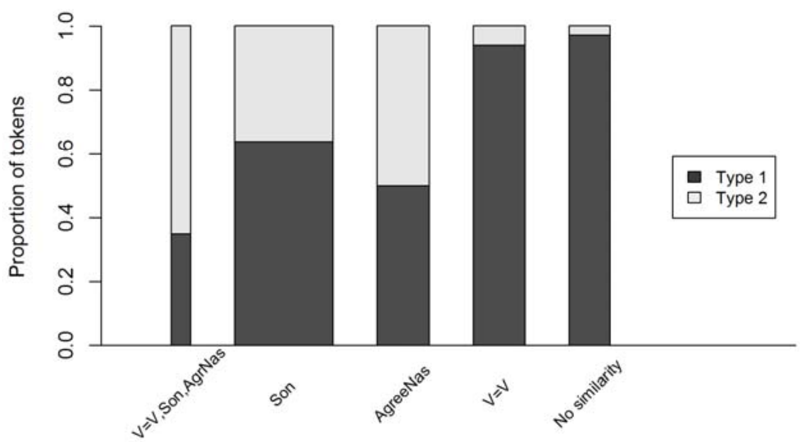
(8) V…V agreement by tone type



2.4 Cumulative similarity

- The more similarity that is exhibited amongst members of the VCV# sequence, the more likely it is for the penultimate vowel to assimilate the definite H tone from the final vowel.
  ↳ amounts to a ganging effect.

(9) Cumulative similarity



3 ABC AND DIOULA TONE HARMONY

- Agreement by Correspondence theory (ABC)
  - consonant harmony (Walker 2000; Hansson 2001; 2010; Rose & Walker 2004)
  - vowel harmony (Sasa 2009; Walker 2009; 2014; Rhodes 2012)
  - dissimilation (e.g., Bennett 2013)
  - local harmonies (e.g., Wayment 2009; Inkelas & Shih 2014; Shih & Inkelas 2014; Lionnet 2014)
  - tone (e.g., Shih 2013; Shih & Inkelas 2014; in prep; Inkelas & Shih 2015)

- ABC analysis for Dioula adapted from Shih 2013.[5]

- Basic ABC insight: surface CORR(espondence) relationships are determined by phonological similarity and proximity of segments. Segments that are similar and proximal will interact. [6]

(10)	CORR-X::X {V}	*Segments with highest amount of sonority (i.e., vowels) correspond.*
	CORR-X::X {V, R}	*Vowels, liquids correspond.*
	CORR-X::X {V, R, N}	*Sonorants (vowels, liquids, nasals) correspond.*
	CORR-X::X	*All segments correspond.*
(11)	CORR-VV [F]	*Vowels identical in feature set* [F] *correspond.*
(12)	CORR-X::X [±nas]	*Segments identical in nasality specification correspond.*[7]

- Unstable surface correspondence occurs when corresponding segments are similar enough to interact but too uncomfortably similar to stably co-exist at a certain distance, as mandated by CORR-LIMITER constraints (e.g., IDENT-XX, XX-EDGE).[8]

(13) IDENT-XX [tone] *Corresponding segments must agree in tone specification.*

- Attraction (i.e., harmony) and repellence (i.e., dissimilation) are repairs for unstable similarity- and proximity-driven surface correspondences (see also Inkelas & Shih 2014 for other repairs triggered by unstable surface correspondence).
  - Relevant CORR constraints and IDENT-XX [tone] must trump input-output faithfulness to tone.

(14) IDENT-IO V[tone] *Maintain input identity of vowel tone specification in the output.*

- Some hand-weighted tableaux to illustrate the system:

(15) Sonorant C_f facilitates tone spread

		weight	4	4	3	1	$\mathcal{H}$
		mèlí	CORR-X::X {V, R}	IDENT-XX [tone]	IDENT-IO V[tone]	CORR-X::X	
	a.	mèlí	-2			-2	-10
	b.	mè$_x$ĺ$_x$í$_x$		-1 (VC)			-4
	c.	mé$_x$l$_x$í$_x$		-2 (VC, CV)	-1		-11
☞	d.	mé$_x$ĺ$_x$í$_x$			-1		-3
	e.	mè$_x$lí$_x$	-2	-1		-2	-14
	f.	mé$_x$lí$_x$	-2		-1	-2	-13

[5] In this talk, I set aside the debates surrounding autosegmentalism versus surface-correspondence based optimization for tonal phenomena; on this topic, see e.g., Zoll 2003; Hyman 2011; 2012; Jurgec 2013; Shih 2013; Inkelas & Shih 2015; Shih & Inkelas, in prep. Similarly, I set aside debates on formalizing agreement for harmony, either via ABC-type constraints, positional licensing, AGREE, etc.
[6] Note: "::" denotes immediately adjacent segments. See Hansson 2001, 2010 for proximity in ABC; notation follows Inkelas & Shih 2014.
[7] Nasality is formulated here as a bivalent feature so that segments agreeing in *orality* also mandate similarity-induced correspondence.
[8] In dealing with consonant dissimilation, Bennett (2013) calls this class of constraints "CC-LIMITER" constraints.

(16) Obstruent C_f does not facilitate tone spread

	weight	4	4	3	1	$\mathcal{H}$
	brìsá	CORR-X::X {V, R}	IDENT-XX [tone]	IDENT-IO V[tone]	CORR-X::X	
☞	a. brìsá				-2	-2
	b. brì$_x$ś$_x$á$_x$		-1 (VC)			-4
	c. brí$_x$s$_x$á$_x$		-2 (VC, CV)	-1		-11
	d. brí$_x$ś$_x$á$_x$			-1		-3
	e. brì$_x$sá$_x$		-1		-2	-6
	f. brí$_x$sá$_x$			-1	-2	-5

(17) Ganging effect via cumulative additivity: sonorant C_f and identical VV's facilitate tone spread

	weight	4	3	4	3	1	$\mathcal{H}$
	tùrú	CORR-X::X {V, R}	CORR-VV [F]	IDENT-XX [tone]	IDENT-IO V[tone]	CORR-X::X	
	a. tùrú	-2	-1			-2	-13
	b. tù$_{xy}$ŕ$_x$ú$_{xy}$			-2 (VC, VV)			-8
	c. tú$_{xy}$r$_x$ú$_{xy}$			-2 (VC, CV)	-1		-11
☞	d. tú$_{xy}$ŕ$_x$ú$_{xy}$				-1		-3
	e. tù$_y$rú$_y$	-2		-1		-2	-14
	f. tú$_y$rú$_y$	-2			-1	-2	-13

- In comparative grammaticality, greater $\Delta\mathcal{H}$ denotes a greater likelihood of tone spread.
- $\Delta\mathcal{H}$ between harmonic winner and completely non-harmonic, non-corresponding candidate (a) = 10
  - cf. $\Delta\mathcal{H}$ for non-identical VV (tableau 15) = 7

(18) Ganging effect via weighted conjunction: sonorant C_f and identical VV's facilitate tone spread

	weight	2	4	3	4	3	1	$\mathcal{H}$
	tùrú	CORR-X::X {V, R} & CORR-VV	CORR-X::X {V, R}	CORR-VV [F]	IDENT-XX [tone]	IDENT-IO V[tone]	CORR-X::X	
	a. tùrú	-2	-2	-1			-2	-17
	b. tù$_{xy}$ŕ$_x$ú$_{xy}$				-2 (VC, VV)			-8
	c. tú$_{xy}$r$_x$ú$_{xy}$				-2 (VC, CV)	-1		-11
☞	d. tú$_{xy}$ŕ$_x$ú$_{xy}$					-1		-3
	e. tù$_y$rú$_y$		-2		-1		-2	-14
	f. tú$_y$rú$_y$		-2			-1	-2	-13

- $\Delta\mathcal{H}$ between harmonic winner and completely non-harmonic, non-corresponding candidate (a) = 14
  - cf. $\Delta\mathcal{H}$ for non-conjoined, additive ganging (tableau 17) = 10

4 MODELING DETAILS

4.1 Maximum Entropy Harmonic Grammar

- Maximum Entropy Harmonic Grammar (MaxEnt) (Goldwater & Johnson 2003; Wilson 2006; Jäger 2007; Hayes & Wilson 2008; a.o.), fitted using MaxEnt Grammar Tool (Hayes, Wilson & George 2009).

- Maximum Entropy models (aka: log-linear models, multinomial logistic regression models) are a general class of statistical models.[9]
  ↳ When there are only two output choices, then MaxEnt = binary logistic regression.
    – fitted here using `bayesglm()` from the `arm R` package (Gelman et al. 2013).

- A point of difference: MaxEnt HG restricts constraint weights to nonnegative numbers, under the assumption that violations of constraints should only penalize rather than reward (though see e.g., Goldrick & Daland 2009 for a possible application of negative weights in MaxEnt; see e.g., Kimper 2011 for an application of constraint rewards for harmony).
    – Regression models don't *a priori* have this restriction. More on this later, if time.

4.2 Weighted conjunction as interaction

- Interactions are already commonplace for linguistic systems, analyses, and beyond (esp. sociolinguistics, psycholingusitics) (e.g., topicality and prototypicality interaction overpowers animacy in English genitive construction choice: Jäger & Rosenbach 2006).
- In regression modeling, interactions are when effect of two constraints on a third are not (merely) additive, and implemented as the product of two constraints, $C_1{*}C_2$, which can receive its own weight w. For example:

(19) $$w_1C_1 + w_2C_2 + w_3(C_1 \times C_2) + \cdots + w_kC_{iN} = \boldsymbol{w_k} \cdot \boldsymbol{C_i} = \mathcal{H}(x,i)$$ (cf. fn. 1)

- When constraint violations ∈{0, 1}, then $C_1{*}C_2 = C_1\ \&\ C_2$, but
- When constraint violations ∈[0, +∞), then $C_1{*}C_2 \neq C_1\ \&\ C_2$.[10]

  ↳ Weighted constraint conjunctions are implemented here as products of two constraints, to remain consistent with statistical regression analyses.

- Instances of weighted interactions in MaxEnt and/or HG:
    – in metrical verse: prosodic position × rhythmic constraints (Hayes et al. 2012)
    – in phonotactics: $[F_1] \times [F_2]$ (Pater & Moreton 2012)
    – in syllable margin phonotactics: $*M_1 \times *M_2$ (Green & Davis 2014)
- Small caveat: for now, setting aside the issue of domain locality in constraint conjunction. See e.g., Łubowicz 2005 for discussion.

[9] Some terminological translations:

Harmonic Grammar ⇔	Statistics
constraint (C_i)	parameter (X_i)
weight (w)	coefficient (β), estimated parameter value ($\hat{\beta}$)
/input form/	conditioning context (x)
candidate set, given by Gen	possible outcomes (Ω or $\mathcal{Y}(x)$)
[output form]	response (y)
weighted constraint conjunction ($C_1{*}C_2$)	interaction term ($X_1{*}X_2$)

[10] where $C_1\&C_2$ denotes the local conjunction of C_1 and C_2, as defined in previous work on LCC.

4.3 Model selection and comparison in weighted OT

- Methodology: compare MaxEnt grammars with and without conjoined constraints.
  - If the gang effect is only additive, then a grammar with conjoined constraints shouldn't contribute any additional explanatory power.
  - If the gang effect is super-cumulative (i.e., not merely additive), then a model with conjoined constraints will demonstrate improved explanatory power over a model without conjoined constraints.

- Model selection and comparison is something rarely talked about in HG/OT (some exceptions in MaxEnt literature: Wilson & Obdeyn 2009; Hayes et al. 2012).
- Usually, the assumption is that learners have a constraint set already provided by CON, and the question is how to fit constraint weights.
  - thus, the assessment (and rejection) of the viability of a constraint's existence is largely left to arguments on conceptual or phonological grounds: e.g., Occam's Razor or learnability arguments.
  - When there is discussion of inductive constraint learning, the induction of conjoined forms of markedness is usually allowed: *[F_1]&[F_2] (e.g., Hayes & Wilson 2008; Moreton 2010).

- Needed: a quantitative way to assess whether the existence of constraint conjunctions significantly improves the grammar.
  - i.e., a way to assess competing conceptions of constraint sets provided by CON.

⇨ **Akaike Information Criterion (AIC) Model Selection and Comparison**

- = approach founded on idea that all models are mere approximations of full reality, an ideal for which the true parameters (β) are unknown (Kullback & Leibler 1951; Anderson & Burnham 2002; Burnham & Anderson 2002; 2004; a.o.).
- AIC-based model selection aims to reduce amount of information loss between candidate models.

- Second-order AIC_c, penalizes for increasing number of constraints against the sample size: [11]

(20) $$AIC_c = -2log\left(\mathcal{L}\left(\hat{\beta}\middle|D\right)\right) + 2K + \frac{2K(K+1)}{n-K-1},$$

where $\mathcal{L}\left(\hat{\beta}\middle|D\right)$ = maximum likelihood of observed data D given fitted parameters $\hat{\beta}$,
K = number of estimable parameters (i.e., constraints) in the model, and
n = sample size.

- Lower AIC_c is better = less information lost; more weight of evidence in favor of the given model.
- In comparing models, $\Delta AIC_c \geq 10$ is considered large:
  - equivalent to ~150 to 1 evidence ratio (E) that the second best model has essentially no support (evidence) of being as good as the best model (Burnham & Anderson 2004:271; Anderson 2008:89–90).[12]

[11] When $n/K > 40$, AIC and AIC_c begin to converge. Because AIC_c regularizes for sample size, it is more conservative to use it for model comparison in general (Burnham & Anderson 2004:269–270).
[12] Evidence ratio, $E_{i,j} = \frac{1}{e^{\left(-(1/2)\Delta_j\right)}}$ for models i and j, where $\Delta_j = AICc_j - AICc_i$.

⇨ **Advantages of AIC$_c$**

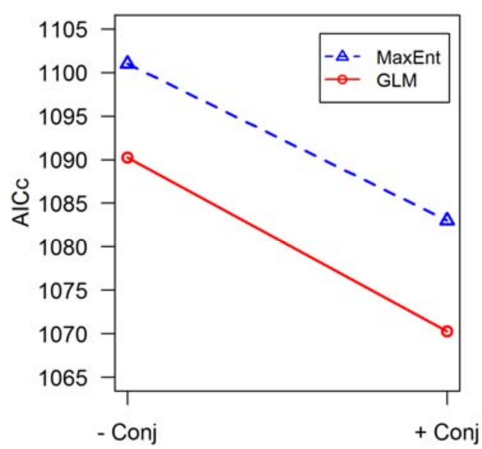
- AIC$_c$ is based on likelihood ratio test and maximizes descriptive accuracy (used by e.g., Hayes et al. 2012), but penalizes for model complexity/loss of restrictiveness in the grammar. [13,14]
- AIC$_c$ can compare weight of evidence for non-nested models over the same data (see also Wilson & Obdeyn 2009). Other tests, like likelihood ratios, cannot.
- AIC$_c$ is computationally fast.
- AIC$_c$ results for assessing restrictiveness and generalisability of models asymptotically converge with *k*-fold cross-validation as the sample size increases (also computationally faster than *k*-fold cross-validation) (Stone 1977; et seq.).

- Warning: AIC$_c$ is **not** a statistical test of significance or stand-alone goodness-of-fit.
  - AIC$_c$ is an information-theoretic *comparison* statistic (see e.g., Anderson 2008 for model comparison philosophy).

5 RESULTS

- Four candidate models for comparison:
  1. MaxEnt, with no constraint conjunction (– Conj)
  2. MaxEnt, with weighted constraint conjunction (+ Conj)
  3. GLM (unbounded coefficients), with no constraint conjunction (– Conj)
  4. GLM (unbounded coefficients), with weighted constraint conjunction (+ Conj)
- Models simplified to binary outcomes: Tone Agreement (via CORRespondence) versus No Tone Agreement (no CORRespondence).

(21) **AIC$_c$ comparison results**



	MaxEnt	**GLM**
– Conj	1101	1090.225
+ Conj	1083	1070.26
ΔAIC$_c$	**18**	**19.965**
E	**8103.08**	**21644.36**

⇨ ΔAIC$_c$ comparisons for both MaxEnt and GLM reveal substantial support that a model with weighted constraint conjunction is a better approximation of truth given the available evidence, even after penalizing for increased model complexity.

---

[13] Wilson & Obdeyn's maximum a posterior (MAP) approach also explicitly penalizes for extreme values of estimated parameters using an assumed prior.

[14] AIC versus Bayesian Information Criterion (BIC): BIC is biased towards models with a few large effects and no other small, tapering effects, which works well for model selection in cases with very large sample sizes (much larger than are normally seen in phonological modeling). AIC, on the other hand, is about bias-variance trade-off and will choose, at a given *n*, a model that is as parsimonious as need to minimize information loss and no more. See e.g., Burnham & Anderson 2004:275ff for detailed discussion. Empirical tests of model selection measures on real-world linguistic data are necessary and encouraged in future work.

5.1 Maximum Entropy results

(22) MaxEnt, – Conj

Constraint	*Weight*
IDENT-IO V[tone]	3.985
CORR-X::X	0.0
CORR-X::X {V, R, N}	1.341
CORR-X::X {V, R}	0.589
CORR-X::X {V}	2.343
CORR-VV [F]	0.187
CORR-X::X [±nas]	0.521

(23) MaxEnt, + Conj

Constraint	*Weight*
IDENT-IO V[tone]	3.588
CORR-X::X	0.0
CORR-X::X {V, R, N}	1.266
CORR-X::X {V, R}	0.118
CORR-X::X {V}	2.516
CORR-VV [F]	0.0
CORR-X::X [±nas]	0.0
CORR-VV [F] * CORR-X::X {V, R, N}	0.0
CORR-VV [F] * CORR-X::X {V, R}	0.263
CORR-VV [F] * CORR-X::X [±nas]	0.259
CORR-X::X [±nas] * CORR-X::X {V, R}	0.466

(24) Comparison of non-AIC model statistics for MaxEnt grammars

	C	D_{xy}		*Likelihood (-2LogLik)*
– Conj	0.8576	0.7152		1085
+ Conj	0.8654	0.7309		1059
			Δ-2LogLik	26 (***, *df*=4, *p* <0.0001)

- Even given standard likelihood ratio test, the model with constraint conjunction is significantly better than one without.
- Testing conjoined constraints reveals that the gang effect of similarity for Dioula is most active for segments that are already highly similar—i.e., between liquids and vowels.
  - e.g., conjunction of CORR-VV [F] * CORR-X::X {V, R, N} receives no weight, but CORR-VV [F] * CORR-X::X {V, R} receives a weight of 0.263.
- Additive effect of conjoined constraints involving liquids is further ganging, at the level of conjoined constraints = super-cumulativity.
  - e.g., a non-agreeing vowel-liquid-vowel sequence such as in **tùrú* would incur the additive violations of not only the simplex constraints but also the conjoined constraints.
  - e.g., model with conjunction assigns *túrú* 71.04% probability versus **tùrú*, 28.96%.
    cf. model without conjunction assigns *túrú* 64.18% probability versus **tùrú*, 35.82%.

5.2 Binomial Generalized Linear Model (non-bounded Maximum Entropy) results[15]

(25) GLM, – Conj

Constraint	*Weight*	*Std. Error*	*z value*	*Pr(>|z|)*
(Intercept)	-4.0756	0.3214	-12.68	<0.0001
CORR-X::X {V, R, N}	1.3144	0.1590	8.27	<0.0001
CORR-X::X {V, R}	0.5766	0.0948	6.08	<0.0001
CORR-X::X {V}	2.4274	0.3707	6.55	<0.0001
CORR-VV [F]	0.4502	0.1606	2.80	0.0051
CORR-X::X [±nas]	0.5097	0.1714	2.97	0.0029

(26) GLM, + Conj

Constraint	*Weight*	*Std. Error*	*z value*	*Pr(>|z|)*
(Intercept)	-3.8181	0.5506	-6.93	<0.0001
CORR-X::X {V, R, N}	1.5524	0.2824	5.50	<0.0001
CORR-X::X {V, R}	0.0501	0.1795	0.28	0.78019
CORR-X::X {V}	2.7530	0.4209	6.54	<0.0001
CORR-VV [F]	0.7478	0.6161	1.21	0.22481
CORR-X::X [±nas]	-0.5937	0.3347	-1.77	0.07608
CORR-VV [F] * CORR-X::X {V, R, N}	-0.5789	0.3228	-1.79	0.07291
CORR-VV [F] * CORR-X::X {V, R}	0.3258	0.1958	1.66	0.09613
CORR-VV [F] * CORR-X::X [±nas]	0.5942	0.3531	1.68	0.09243
CORR-X::X [±nas] * CORR-X::X {V, R}	0.6443	0.1816	3.55	0.00039

(27) Comparison of non-AIC model statistics for GLM models

	C	D_{xy}	R^2		*Likelihood (-2LogLik)*
– Conj	0.8399	0.6797	0.4939		1078.155
+ Conj	0.8479	0.6959	0.5199		1050.074
				Δ-2LogLik	28 (***, *df*=4, p <0.0001)

A brief tangent:
- A significant point of difference between MaxEnt HG and logistic regression models is the general restriction in HG of weights to non-negative values (i.e., zero-bounded).
- Negative estimated parameter values in GLM model with conjunctions = rewards for *not* tonally agreeing when segments are insufficiently similar (cf. Kimper 2011 for rewards in vowel harmony).
  - e.g., negative estimated value for CORR-X::X [±nas] (β = -0.5937) rewards the lack of tone agreement (and CORRespondence) between nasals, obstruents, and their surrounding vowels, as compared to penalizing the lack of tone agreement between liquids and their surrounding vowels (CORR-X::X [±nas] * CORR-X::X {V, R}, β = 0.6443).
  - e.g., negative estimated value for conjunction of CORR-VV [F] * CORR-X::X {V, R, N} indicates possible *disjunction* term (cf. Crowhurst & Hewitt 1997), where nasals are less likely to have a similarity ganging effect with the surrounding vowels than liquids.
- Negative values highlight similarity bases of ganging effect for tone harmony.
- Negative values for interactions also allow for evaluation of constraint *disjunctions* (see e.g., Crowhurst & Hewitt 1997).

[15] In the following models, the intercept stands in for IDENT-IO V[tone]. I.e., default is no tone agreement.

6 DISCUSSION & CONCLUSION

- Arguments against constraint conjunction are typically of two kinds:
  1. Theoretical parsimony, and
  2. Learnability of constraint conjunction

⇨ **Theoretical parsimony**

- Opponents of conjunction argue that constraint conjunctions should be *a priori* restricted from CON so as to maintain restrictiveness (i.e., reduce complexity) in the constraint space (e.g., Potts et al. 2010; Pater 2015).
- But the argument of theoretical parsimony cuts both ways: we can maintain restrictiveness in our basic theoretical assumptions by allowing an unrestricted constraint space and letting the grammar do the choosing of relevant constraints.
- **Information-theoretic model comparison method** presented here gets at the best of both worlds: allow only the constraint conjunctions that are shown to improve the model, despite a penalization for increased model complexity.[16]
  - ↳ Super-cumulativity may be lurking in your data.

⇨ **Learnability of constraint conjunction**

- Opponents of conjunction argue that there is no learnability theory for constraint conjunction.
  - – How are constraint conjunctions chosen/induced?
  - – How are constraint conjunctions limited? (i.e., why not crazy conjunctions of every possible constraint combination?)
- vs. Additive cumulativity in HG, which solves at least some of these problems:
  - – Additive cumulativity comes for free as part of the HG package, which has an associated, proven learning algorithm.

- **Combining ABC with MaxEnt** for similarity-based harmony, at least, conjunctions could be phonologically-grounded in feature theory: see e.g., Wayment 2009 on phonological similarity attraction as conjunction/ganging.
- **Weighted constraint conjunction** provides the theory of conjunction a fair chance to be evaluated in a probabilistic model. Previous comparisons of conjunction and HG were confounded by comparing only strict-ranking OT+conjunction versus HG–conjunction.
- A future proposal for learning constraint conjunctions: weighted constraint conjunction potentially draws from additive cumulativity in HG as part of learning.
- If a learner sees enough weight of evidence that there are cumulative effects from additive constraint interactions, then a separate and independent conjunction of those constraints is posited, reducing the extreme values of simplex constraints in favor of a grammar with justifiably more complex parameters and better accuracy.
  - – Potential advantage: only conjunctions that will be posited are ones that utilize constraints already in an additively cumulative interaction.
  - ↳ Additive cumulativity guides learning of weighted constraint conjunction for super-cumulativity effects.

Acknowledgements to Kie Zuraw, Brian Smith, Laura McPherson, Sharon Inkelas, Gunnar Hansson, Robert Daland, and audiences at WCCFL 33 and UCs Merced and Berkeley for discussion.

[16] Notion of "model complexity" could encompass many aspects, not limited to increased numbers of parameters. E.g., Wilson & Obdeyn 2009 argue that extreme values of parameters should also be penalized as a mark of model complexity.

7 References

Albright, Adam. 2009. Cumulative violations and complexity threshold: Evidence from Lakhota. Paper. Paper presented at the 17th Manchester Phonology Meeting, Manchester, UK.

Anderson, David R. 2008. *Model Based Inference in the Life Sciences: A Primer on Evidence*. New York, NY: Springer.

Anderson, David R. & Kenneth P. Burnham. 2002. Avoiding pitfalls when using information-theoretic methods. *Journal of Wildlife Management* 66(3). 912–918.

Baković, Eric. 2000. Harmony, dominance, and control. Rutgers University Ph.D. dissertation.

Bennett, William G. 2013. Dissimilation, consonant harmony, and surface correspondence. New Brunswick, NJ: Rutgers, The State University of New Jersey Ph.D. dissertation.

Braconnier, Cassian. 1982. *Le système tonal du dioula d'Odienné*. Abidjan, Ivory Coast: Université d'Abidjan.

Braconnier, Cassian & S Diaby. 1982. *Dioula d'Odienné: Material Lexical*. (96). Institut de Linguistique Appliquee: Agence de cooperation culturelle et technique.

Burnham, Kenneth P. & David R. Anderson. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. 2nd edition. New York, NY: Springer-Verlag.

Burnham, Kenneth P. & David R. Anderson. 2004. Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research* 33(2). 261–304. doi:10.1177/0049124104268644.

Crowhurst, Megan & Mark Hewitt. 1997. *Boolean operations and constraint interaction in Optimality Theory*. Chapel Hill, NC, and Waltham, Mass.

Farris-Trimble, Ashley W. 2008. Cumulative faithfulness effects in phonology. Indiana University Ph.D. dissertation.

Gelman, Andrew, Yu-Sung Su, Masanao Yajima, Jennifer Hill, Maria Grazia Pittau, Jouni Kerman & Tian Zheng. 2013. *Package "arm."* http://CRAN.R-project.org/package=arm.

Goldrick, Matthew & Robert Daland. 2009. Linking speech errors and phonological grammars: insights from Harmonic Grammar networks. *Phonology* 26(1). 147–185.

Goldwater, Sharon & Mark Johnson. 2003. Learning OT Constraint Rankings Using a Maximum Entropy Model. In Jennifer Spenader, Anders Eriksson & Östen Dahl (eds.), *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, 111–120. Stockholm: Stockholm University.

Green, Christopher & Stuart Davis. 2014. Superadditivity and Limitations on Syllable Complexity in Bambara Words. In Ashley W. Farris-Trimble & Jessica Barlow (eds.), *Perspectives on Phonological Theory and Acquisition: Papers in Honor of Daniel Dinnsen*. Philadelphia/Amsterdam: John Benjamins Co.

Hansson, Gunnar Ólafur. 2001. Theoretical and Typological Issues in Consonant Harmony. University of California, Berkeley Ph.D. dissertation.

Hansson, Gunnar Ólafur. 2010. *Consonant Harmony: Long-Distance Interaction in Phonology*. (UC Publications in Linguistics). Berkeley, CA: University of California Press.

Hayes, Bruce & Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39(3). 379–440.

Hayes, Bruce, Colin Wilson & Ben George. 2009. *Maxent Grammar Tool*. http://www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool/.

Hayes, Bruce, Colin Wilson & Anne Shisko. 2012. Maxent grammars for the metrics of Shakespeare and Milton. *Language* 88(4). 691–731.

Hyman, Larry M. 2011. Tone: Is it different? In John Goldsmith, Jason Riggle & Alan C. L. Yu (eds.), *The Handbook of Phonological Theory*, 197–239. 2nd ed. Oxford, UK: Blackwell Publishing.

Hyman, Larry M. 2012. How autosegmental is phonology? Paper. Paper presented at the Colloque du Réseau Français de Phonologie, Paris, France.

Inkelas, Sharon & Stephanie S Shih. 2014. Unstable surface correspondence as the source of local conspiracies. In Jyoti Iyer & Leland Kusmer (eds.), *NELS 44: Proceedings of the 44th Meeting of the North East Linguistic Society*, vol. 1, 191–204. University of Massachusetts, Amherst: GLSA Publications.

Inkelas, Sharon & Stephanie S Shih. 2015. Tone melodies in the age of Surface Correspondence. Paper. Paper presented at the 51st Meeting of the Chicago Linguistic Society (CLS), University of Chicago.

Ito, Junko & Armin Mester. 2003. *Japanese Morphophonemics: Markedness and Word Structure*. Cambridge, Massachusetts: MIT Press Linguistic Inquiry Monograph Series 41.

Jäger, Gerhard. 2007. Maximum Entropy Models and Stochastic Optimality Theory. In Jane Grimshaw, Joan Maling, Chris Manning, Jane Simpson & Annie Zaenen (eds.), *Architectures, Rules, and Preferences: A Festschrift for Joan Bresnan*. Stanford, CA: CSLI.

Jäger, Gerhard & Anette Rosenbach. 2006. The winner takes it all—almost. *Linguistics* 44. 937–971.

Jurgec, Peter. 2013. Two types of parasitic assimilation. (Ed.) Sylvia Blaho, Martin Krämer & Bruce Morén-Duolljá. *Nordlyd* 40(1). (Special Issue "A Festschrift on the Occasion of X Years of CASTL Phonology and Curt Rice's Lth Birthday"). 108–135.

Kimper, Wendell A. 2011. Competing triggers: Transparency and opacity in vowel harmony. Amherst, MA: University of Massachusetts, Amherst Ph.D. dissertation.

Kullback, Soloman & Richard A. Leibler. 1951. On Information and Sufficiency. *Annals of Mathematical Studies* 2. 79–86.

Legendre, Geraldine, Yoshiro Miyata & Paul Smolensky. 1990. Harmonic Grammar – A formal multi-level connectionist theory of linguistic well-formedness: An Application. *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*, 884–891. Mahwah, NJ: Lawrence Erlbaum Associates.

Legendre, Géraldine, Antonella Sorace & Paul Smolensky. 2006. The Optimality Theory–Harmonic Grammar Connection. In Paul Smolensky & Géraldine Legendre (eds.), *The Harmonic Mind*, 339–402. Cambridge, MA: MIT Press.

Lionnet, Florian. 2014. Doubly Triggered Harmony in Laal as Subphonemic Agreement by Correspondence. In John Kingston, Claire Moore-Cantwell, Joe Pater & Robert Staubs (eds.), *Supplemental Proceedings of the 2013 Meeting on Phonology*. Washington, DC: Linguistic Society of America. http://journals.linguisticsociety.org/proceedings/index.php/amphonology/issue/view/1.

Lubowicz, Anna. 2005. Locality of Conjunction. *Proceedings of WCCFL 24*. Somerville, MA: Cascadilla. http://www.lingref.com/cpp/wccfl/24/paper1230.pdf.

Moreton, Elliott. 2010. Constraint induction and simplicity bias in phonotactic learning. Paper. Paper presented at the Workshop on Grammar Induction, Cornell University. http://www.unc.edu/~moreton/Papers/MoretonCornellWkshp2010.pdf.

Pater, Joe. 2015. Universal Grammar with Weighted Constraints. In John J McCarthy & Joe Pater (eds.), *Harmonic Grammar and Harmonic Serialism*. London: Equinox Press.

Pater, Joe & Elliot Moreton. 2012. Structurally biased phonology: Complexity in learning and typology. *The English and Foreign Languages Journal* 3(2). 1–44.

Potts, Christopher, Joe Pater, Karen Jesney, Rajesh Bhatt & Michael Becker. 2010. Harmonic Grammar with linear programming: from linear systems to linguistic typology. *Phonology* 27(01). 77–117. doi:10.1017/S0952675710000047.

Rhodes, Russell. 2012. *Vowel Harmony as Agreement by Correspondence*. (UC Berkeley Phonology Lab Annual Report). University of California, Berkeley. http://linguistics.berkeley.edu/phonlab/annual_report/annual_report_2012.html.

Rose, Sharon & Rachel Walker. 2004. A Typology of Consonant Agreement by Correspondence. *Language* 80(3). 475–531.

Sasa, Tomomasa. 2009. Treatment of vowel harmony in Optimality Theory. Iowa City, IA: University of Iowa Ph.D. dissertation.

Shih, Stephanie S. 2013. Consonant-tone interaction as Agreement by Correspondence. Ms. Stanford University and University of California, Berkeley, ms. http://stanford.edu/~stephsus/shih-ctoneABC-draftms_1-18-13.pdf.

Shih, Stephanie S & Sharon Inkelas. in prep. Autosegmental Aims in Surface Optimizing Phonology. University of California. Merced and Berkeley, ms.

Shih, Stephanie S & Sharon Inkelas. 2014. A subsegmental correspondence approach to contour tone (dis)harmony patterns. *Proceedings of Phonology 2013*. University of Massachusetts, Amherst. http://journals.linguisticsociety.org/proceedings/index.php/amphonology/article/view/22.

Smolensky, Paul. 1993. Harmony, markedness, and phonological activity. New Brunswick, NJ.

Smolensky, Paul. 2006. Optimality in Phonology II: Harmonic Completeness, Local Constraint Conjunction, and Feature Domain Markedness. *The Harmonic Mind*, 27–160. MIT Press.

Smolensky, Paul & Géraldine Legendre. 2006. *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar*. Cambridge, MA: MIT Press.

Stone, M. 1977. An Aymptotic Equivalence of Choice of Model by Cross-Validation and Akaike's Criterion. *Journal of the Royal Statistical Society. Series B (Methodological)* 39(1). 44–47.

Walker, Rachel. 2000. Long-distance Consonantal Identity Effects. In Roger Billerey & Brook Danielle Lillehaugen (eds.), *Proceedings of WCCFL 19*, 532–545. Somerville, MA: Cascadilla Press.

Walker, Rachel. 2009. Similarity-sensitive blocking and transparency in Menominee. Paper. Paper presented at the 83rd Annual Meeting of the Linguistic Society of America, San Francisco, California. http://www-bcf.usc.edu/~rwalker/MenomineeLSAHdt.pdf.

Walker, Rachel. 2014. Prominence-control and Multiple Triggers in Vowel Harmony: An ABC Analysis. *ABC<->Conference Archive*, vol. UC Berkeley Phonology Lab Annual Report, 202–213. University of California, Berkeley. http://linguistics.berkeley.edu/phonlab/annual_report/documents/2014/ABCC/Walker.pdf.

Wayment, Adam. 2009. Assimilation as attraction: Computing distance, similarity, and locality in phonology. Baltimore, MD: Johns Hopkins University Ph.D. dissertation.

Wilson, Colin. 2006. Learning Phonology with Substantive Bias: An Experimental and Computational Study of Velar Palatalization. *Cognitive Science* 30(5). 945–982.

Wilson, Colin & Marieke Obdeyn. 2009. Simplifying subsidiary theory: statistical evidence from Arabic, Muna, Shona, and Wargamay.

Zoll, Cheryl. 2003. Optimal Tone Mapping. *Linguistic Inquiry* 34(2). 225–268.